Tutorial 11: Kernel Methods

Polynomial features, RBF geometry, and kernel SVM

Kangning Cui
City University of Hong Kong (Dongguan)

Semester A, 2026/27

Task 1: Polynomial kernel

Consider three scalar observations:

$$x_1 = -1, \qquad x_2 = 0, \qquad x_3 = 2.$$

Use the polynomial kernel and its feature map:

$$k(x, z) = (1 + xz)^2, \qquad \phi(x) = \begin{bmatrix} 1 \\ \sqrt{2}x \\ x^2 \end{bmatrix}.$$

The Gram matrix is defined by $K_{ij} = k(x_i, x_j)$.

(a) Compute the 3×3 Gram matrix $\boldsymbol{K}$.

(b) Find $\phi(x_1)$ and $\phi(x_3)$, and verify $\phi(x_1)^\mathsf{T}\phi(x_3) = K_{13}$.

(c) Compute the squared feature-space distance $\|\phi(x_1) - \phi(x_3)\|_2^2$ from kernel values. Compare it with $(x_1 - x_3)^2$.

(d) For $s(x, z) = k(x, z)/\sqrt{k(x, x)k(z, z)}$, compute $s(x_1, x_2)$ and $s(x_1, x_3)$. Identify the pair with greater normalized similarity.

Solution 1(a): Gram matrix

The six distinct kernel values are:

$$\begin{aligned} k(-1,-1) &= (1+1)^2 = 4, & k(-1,0) &= 1, \\ k(-1,2) &= (1-2)^2 = 1, & k(0,0) &= 1, \\ k(0,2) &= 1, & k(2,2) &= (1+4)^2 = 25. \end{aligned}$$

By symmetry, $K_{ij} = K_{ji}$. Therefore:

$$\boxed{\boldsymbol{K} = \begin{bmatrix} 4 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 25 \end{bmatrix}.}$$

Every entry is an inner product of the corresponding feature vectors.

Solution 1(b): Explicit features

Evaluate the feature map at $x_1 = -1$ and $x_3 = 2$:

$$\phi(-1) = \begin{bmatrix} 1 \\ -\sqrt{2} \\ 1 \end{bmatrix}, \qquad \phi(2) = \begin{bmatrix} 1 \\ 2\sqrt{2} \\ 4 \end{bmatrix}.$$

The explicit feature-space inner product is:

$$\phi(-1)^{\mathsf{T}}\phi(2) = 1 - 4 + 4 = \boxed{1}.$$

Using the kernel gives exactly the same result:

$$\boxed{k(-1, 2) = (1 - 2)^2 = 1 = K_{13}.}$$

The kernel avoids constructing feature vectors for this calculation.

Solution 1(c): Feature-space distance

Expand the squared distance into kernel values:

$$\begin{aligned}\|\phi(x_1) - \phi(x_3)\|_2^2 &= k(x_1, x_1) + k(x_3, x_3) - 2k(x_1, x_3) \\ &= 4 + 25 - 2(1) = \boxed{27}.\end{aligned}$$

The original squared distance is:

$$(x_1 - x_3)^2 = (-1 - 2)^2 = \boxed{9}.$$

The polynomial map changes the distance used by a learning algorithm.

Solution 1(d): Normalized similarity

Use the kernel values from part (a):

$$K_{11} = 4, \quad K_{22} = 1, \quad K_{33} = 25, \quad K_{12} = K_{13} = 1.$$

Then:

$$s(x_1, x_2) = \frac{1}{\sqrt{4 \cdot 1}} = \boxed{\frac{1}{2}},$$

$$s(x_1, x_3) = \frac{1}{\sqrt{4 \cdot 25}} = \boxed{\frac{1}{10}}.$$

Thus x_1 and x_2 have greater normalized similarity, although their raw kernel value equals $k(x_1, x_3) = 1$.

Task 2: RBF geometry

Consider scalar observations $x_A = 0$, $x_B = 1$, and $x_C = 3$. Use the RBF kernel:

$$k_\gamma(x, z) = e^{-\gamma(x-z)^2}, \qquad \gamma = \tfrac{1}{2}.$$

You may use $e^{-1/2} \approx 0.6065$, $e^{-2} \approx 0.1353$, and $e^{-9/2} \approx 0.0111$.

(a) Compute $k_\gamma(x_A, x_B)$, $k_\gamma(x_B, x_C)$, and $k_\gamma(x_A, x_C)$.

(b) Compute the squared feature-space distances $d^2_\phi(x_A, x_B)$ and $d^2_\phi(x_A, x_C)$, where $d^2_\phi(x, z) = 2 - 2k_\gamma(x, z)$. Compare the distances.

(c) Change γ to 2. Recompute $k_\gamma(x_A, x_B)$ and $d^2_\phi(x_A, x_B)$.

(d) Explain the effect of increasing γ on RBF similarity and the flexibility of an RBF SVM, with C fixed.

Solution 2(a): RBF similarities

Substitute the squared input distances into the kernel:

$$k_{1/2}(0,1) = e^{-(1/2)(1)^2} = \boxed{e^{-1/2} \approx 0.6065},$$
$$k_{1/2}(1,3) = e^{-(1/2)(2)^2} = \boxed{e^{-2} \approx 0.1353},$$
$$k_{1/2}(0,3) = e^{-(1/2)(3)^2} = \boxed{e^{-9/2} \approx 0.0111}.$$

Their input distances are 1, 2, and 3, respectively.

For fixed $\gamma > 0$, closer points have greater RBF similarity.

Solution 2(b): Feature-space distances

For an RBF kernel, $k_\gamma(x, x) = 1$, so:

$$d_\phi^2(x, z) = 2 - 2k_\gamma(x, z).$$

Therefore:

$$d_\phi^2(x_A, x_B) = 2 - 2e^{-1/2} \approx \boxed{0.7870},$$
$$d_\phi^2(x_A, x_C) = 2 - 2e^{-9/2} \approx \boxed{1.9778}.$$

The second pair is farther apart in feature space.

Both squared feature-space distances are below 2, despite the input-space squared distances being 1 and 9.

Solution 2(c): Kernel scale

For the same pair $x_A = 0$, $x_B = 1$, set $\gamma = 2$:

$$\boxed{k_2(0,1) = e^{-2} \approx 0.1353.}$$

Its new squared feature-space distance is:

$$\begin{aligned} d^2_{\phi,\gamma=2}(0,1) &= 2 - 2e^{-2} \\ &\approx \boxed{1.7294}. \end{aligned}$$

At $\gamma = 1/2$, the same values were 0.6065 and 0.7870.

Increasing γ makes this pair less similar and farther apart in the feature-space metric.

Solution 2(d): Locality

For two distinct inputs at fixed squared distance d^2:

$$k_\gamma(x, z) = e^{-\gamma d^2}.$$

Increasing γ makes the kernel value decrease more rapidly with distance.

$$\boxed{\gamma \uparrow \implies \text{more local similarity.}}$$

At fixed C, an RBF SVM can generally form more detailed decision boundaries with larger γ.

If γ is too large, the model may become sensitive to individual training observations or noise.

Task 3: Kernel SVM prediction

A trained soft-margin SVM ($C = 3$) uses the RBF kernel $k(x, z) = 2^{-(x-z)^2}$ and has bias $b = 0$.

i	1	2	3	4
x_i	0	1	$\frac{3}{2}$	$-\frac{1}{2}$
y_i	-1	$+1$	$+1$	-1
α_i	2	2	0	0

Its score is $f(x) = \sum_{i=1}^{4} \alpha_i y_i k(x_i, x) + b$. You may use $2^{-1/4} \approx 0.8409$.

(a) Verify $\sum_i \alpha_i y_i = 0$. Identify the support vectors and compute $f(x_1)$ and $f(x_2)$.

(b) Compute $f(-1)$ and predict the class of $x = -1$.

(c) Compute $f(3)$ and predict the class of $x = 3$.

(d) Compute the signed margins $m_i = y_i f(x_i)$ and hinge losses $\max(0, 1 - m_i)$ for $i = 3, 4$. Relate them to $\alpha_3 = \alpha_4 = 0$.

Solution 3(a): Support vectors

The SVM equality constraint holds:

$$\sum_{i=1}^{4} \alpha_i y_i = 2(-1) + 2(+1) + 0 + 0 = \boxed{0}.$$

Only α_1 and α_2 are positive, so the support vectors are $\boxed{x_1 = 0,\ x_2 = 1}$.

$$f(0) = -2k(0,0) + 2k(1,0) = -2 + 2(\tfrac{1}{2}) = \boxed{-1},$$
$$f(1) = -2k(0,1) + 2k(1,1) = -2(\tfrac{1}{2}) + 2 = \boxed{+1}.$$

Both support vectors lie on the canonical margin: $y_1 f(0) = y_2 f(1) = 1$.

Solution 3(b): Prediction at -1

Only the two support vectors contribute:

$$k(0,-1) = 2^{-1} = \tfrac{1}{2},$$
$$k(1,-1) = 2^{-4} = \tfrac{1}{16}.$$

The decision score is:

$$f(-1) = -2\left(\tfrac{1}{2}\right) + 2\left(\tfrac{1}{16}\right)$$
$$= -1 + \tfrac{1}{8} = \boxed{-\tfrac{7}{8}}.$$

Since $f(-1) < 0$, the prediction is $\boxed{\hat{y} = -1}$.

Solution 3(c): Prediction at 3

For the new input $x = 3$:

$$k(0,3) = 2^{-9} = \tfrac{1}{512},$$
$$k(1,3) = 2^{-4} = \tfrac{1}{16}.$$

Therefore:

$$f(3) = -2\left(\tfrac{1}{512}\right) + 2\left(\tfrac{1}{16}\right)$$
$$= -\tfrac{1}{256} + \tfrac{1}{8} = \boxed{\tfrac{31}{256}}.$$

Since $f(3) > 0$, the prediction is $\boxed{\hat{y} = +1}$.

Its score magnitude is smaller because the point is farther from the support vectors.

Solution 3(d): Margins and hinge loss

For $x_3 = 3/2$:

$$m_3 = y_3 f(\tfrac{3}{2}) = -2 \cdot 2^{-9/4} + 2 \cdot 2^{-1/4}$$
$$= \tfrac{3}{2}\, 2^{-1/4} \approx \boxed{1.261}.$$

By symmetry, $m_4 = y_4 f(-1/2) = m_3$.

$$\boxed{\ell_3 = \ell_4 = \max(0, 1 - m_3) = 0.}$$

Both points lie outside the unit margin. Their zero multipliers are consistent with the SVM KKT conditions.

Kernel computations and SVM decisions

Questions and discussion